控制论 tutorial

甄景贤 (King-Yin Yan)

General.Intelligence@Gmail.com

1 控制论

以下内容可以在一般「现代控制论」教科书中找到，例如：

- Daniel Liberzon 2012: *Calculus of variations and optimal control theory – a concise introduction*
- 李国勇 2008：《最优控制理论与应用》
- 张洪钺、王青 2005：《最优控制理论与应用》

一个动态系统 (dynamical system) 可以用以下方法定义：

$$\text{离散时间：} \quad \boldsymbol{x}_{t+1} = \boldsymbol{F}(\boldsymbol{x}_t) \tag{1}$$

$$\text{连续时间：} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \tag{2}$$

其中 f 也可以随时间改变。如果 f 不依赖时间，则系统是 time-invariant （定常的），形式上如 (2) 那种微分方程叫作 autonomous （自主的）。

在我的智能系统理论里，我把 F 或 f 设定成 RNN (recurrent neural network)，即反馈式神经网络：

$$\text{离散时间：} \quad \boldsymbol{x}_{t+1} = \boxed{\text{RNN}}(\boldsymbol{x}_t) \tag{3}$$

$$\text{连续时间：} \quad \dot{\boldsymbol{x}} = \boxed{\text{RNN}}(\boldsymbol{x}) \tag{4}$$

这里 recurrent 指的是它不断重复作用在 $\boldsymbol{x}$ 之上，但实际上它是一个普通的前馈式 (feed-forward) 神经网络。注意：在抽象理论中，f 和 F 可以是任意函数，我把它们设计成 NN 只是众多可能的想法之一。之所以选用 NN，是因为它有 universal function approximator 的功能，而且是我们所知的最「聪明」的学习机器之一。

在我提出的智能系统里，$\dot{x}$ 是由學習機器給出的，換句話說，$\dot{x}$ 是思維狀態在梯度下降至最佳狀態時的方向導數。

一个（连续时间的）控制系统 (control system) 定义为：

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t), t) \tag{5}$$

其中 $\boldsymbol{u}(t)$ 是控制向量。控制论的目的就是找出最好的 $\boldsymbol{u}(t)$ 函数，令系统由初始状态 $\boldsymbol{x}_0$ 去到终点状态 $\boldsymbol{x}_\perp$。

注意：人工智能中的 **A* search**，是动态规划的一个特例。换句话说，用动态规划在某个空间中「漫游」，可以模拟到 best-first 搜寻的功能。

在这框架下，智能系统的运作可以分开成两方面：思考 和 学习。

思考即是根据已学得的知识（知识储存在 RNN 里），在思维空间中找寻 $\boldsymbol{x}$ 最优的轨迹，方法是用控制论计算 $\boldsymbol{u}^*$。$\boldsymbol{x}$ 的轨迹受 RNN 约束（系统只能依据「正确」的知识去思考），但思考时 RNN 是不变的。

学习就是学习神经网络 RNN 的 weights W。此时令 $u = 0$，即忽略控制论方面。

以上两者是两个独立的方面，但不排除它们可以在实际中同时进行。

1.1 控制论与强化学习的关系

在强化学习中，我们关注两个数量：

- $R(\boldsymbol{x}, a)$ = 在状态 $\boldsymbol{x}$ 做动作 a 所获得的奖励(reward)
- $U(\boldsymbol{x})$ = 状态 $\boldsymbol{x}$ 的效用(utility) 或 价值 (value)

简单来说，「价值」就是每个瞬时「奖励」对时间的积分：

$$\boxed{\text{价值 U}} = \int \boxed{\text{奖励 R}}\, dt \tag{6}$$

（价值有时用 V 表示，但为避免和势能 V 混淆故不用。）

用控制论的术语，通常定义 cost functional：

$$J = \int L dt + \Phi(\boldsymbol{x}_\perp) \tag{7}$$

其中 L 是 “running cost”，即行走每一步的「价钱」；Φ 是 terminal cost，即到达终点 $\boldsymbol{x}_\perp$ 时，那位置的价值。

在分析力学里 L 又叫 Lagrangian，而 L 对时间的积分叫「作用量」：

$$\boxed{\text{作用量 (Action)}}\ S = \int L dt \tag{8}$$

Hamilton 的最小作用量原理 (principle of least action) 说，在自然界的运动轨迹里，S 的值总是取稳定值 (stationary value)，即比起邻近的轨迹它的 S 值最小。

所以有这些对应：

强化学习	最优控制	分析力学
效用/价值 U	价钱 J	作用量 S
即时奖励 R	running cost	Lagrangian L
action a	control u	（外力?）

用比较浅显的例子：和美女做爱能带来即时的快感 (= 奖励 R)，但如果强奸的话会坐牢，之后很长时间很苦闷，所以这个做法的长远价值 U 比其他做法较低，正常人不会选择它。

有趣的是，奖励 R 对应於力学上的 Lagrangian，其物理学单位是「能量」；换句话说，「快感」或「开心」似乎可以用「能量」的单位来量度，这和通俗心理学里常说的「正能量」不谋而合。而，长远的价值，是以 [能量 $\times$ 时间] 的单位来量度。

一个智能系统，它有「智慧」的条件，就是每时每刻都不断追求「开心能量」或奖励 R 的最大值，但它必需权衡轻重，有计划地找到长远的效用 U 的最大值。

1.2 经典分析力学 (analytical mechanics)

分析力学的物理内容，完全是牛顿力学的 $F = ma$，但在表述上引入了能量和 Hamiltonian 等概念，再使用微积分和变分法。

Lagrange 方程

Lagrange 引入了 Lagrangian $L = T - V$，可以分拆成動能 T 和勢能 V 兩部分。

重點是：動能 T 是速度 $\dot{\boldsymbol{x}}$ 的函數，勢能 V 是位置 $\boldsymbol{x}$ 的函數。

问题：如果在强化学习中的「快感 / 奖励」对应於 Lagrangian L，如何在奖励之中分拆出「动能」和「势能」的分量？

$$\boxed{\text{Lagrange equation}} \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} - \frac{\partial L}{\partial x_i} = 0 \tag{9}$$

这些方程的座标是 $(\boldsymbol{x}, \dot{\boldsymbol{x}})$，可以了解成位置空间 (configuration space)上的 tangent bundle（下述）。

Hamilton 方程

Hamiltonian $H = T + V$，亦即总能量，但它表示成位置 $\boldsymbol{x}$ 和动量 $\boldsymbol{p}$ 的函数。

$$\boxed{\text{Hamilton equation}} \quad \begin{cases} \dot{\boldsymbol{x}} &= \dfrac{\partial H}{\partial p} \\ \dot{\boldsymbol{p}} &= -\dfrac{\partial H}{\partial x} \end{cases} \tag{10}$$

这些方程的座标是相位空间 (phase space) $(\boldsymbol{x}, \boldsymbol{p})$。

位置空间和相位空间之间的变换是 **Legendre transformation**：

$$\boxed{\text{tangent bundle}} \; TX \to T^*X \; \boxed{\text{cotangent bundle}} \tag{11}$$

$$(\boldsymbol{x}, \dot{\boldsymbol{x}}) \mapsto (\boldsymbol{x}, \boldsymbol{p}) \tag{12}$$

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \quad \Rightarrow \quad H := \boldsymbol{p}\dot{\boldsymbol{x}} - L \tag{13}$$

Hamilton-Jacobi 方程

$$\boxed{\text{Hamilton-Jacobi equation}} \quad H(q, \frac{\partial S}{\partial q}, t) + \frac{\partial S}{\partial t} = 0 \tag{14}$$

其中 S 是「作用量」。下面我们会用动态规划的原理推导出此一方程。

Poisson 括号

$$\boxed{\text{Poisson 括号}} \quad \{F, H\} := \sum_i \{\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q_i}\} \tag{15}$$

在力学系统中，它表示任意一力学量（函数 f）对时间的改变量：

$$\dot{f} = \{f, H\} \tag{16}$$

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial p}\dot{p} + \frac{\partial}{\partial q}\dot{q} \tag{17}$$

以上使用了微分的 chain rule，但由於 $\dot{\boldsymbol{p}}$ 和 $\dot{\boldsymbol{q}}$ 可以由 Hamilton 方程 (10) 给出，所以得到 Poisson 括号的形式 (15)。

每个物理学生都知道的「经典－量子对应原理」：

$$[\boldsymbol{F}, \boldsymbol{G}] \quad \Leftrightarrow \quad i\hbar\{F, G\} \tag{18}$$

这对应原理是 P.A.M. Dirac 发现的。

在经典力学里，

$$\{\boldsymbol{x}, \boldsymbol{p}\} = 1 \tag{19}$$

但在量子力学里，

$$[\boldsymbol{X}, \boldsymbol{P}] = i\hbar \tag{20}$$

这也是 Heisenberg 测不准原理的由来：

$$\Delta\boldsymbol{X}\Delta\boldsymbol{P} \geq \frac{\hbar}{2} \tag{21}$$

如果在某一流形上，「广义」Poisson 括号是 nondegenerate（非退化）的，则它变成了辛流形结构的 $\omega = \{\cdot, \cdot\}$ 括号（下述）。

1.3 Hamiltonian 的出现

考虑一个典型的控制论问题，系统是：

$$\text{状态方程：}\quad \dot{\boldsymbol{x}}(t) = \boldsymbol{f}[\boldsymbol{x}(t), \boldsymbol{u}(t), t] \tag{22}$$

$$\text{边值条件：}\quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0\,,\ \boldsymbol{x}(t_\perp) = \boldsymbol{x}_\perp \tag{23}$$

$$\text{目標函数：}\quad J = \int_{t_0}^{t_\perp} L[\boldsymbol{x}(t), \boldsymbol{u}(t), t] dt \tag{24}$$

要找的是最优控制 $\boldsymbol{u}^*(t)$。

> Lagrange multiplier 是找极大/小值的常用方法：如果我们要找：
>
> $$\max\ f(x) \quad \text{subject to} \quad g(x) = 0 \tag{25}$$
>
> Lagrange 建议我们建构 Lagrangian 函数：
>
> $$L(x, \lambda) = f(x) - \lambda g(x) \tag{26}$$
>
> 然后求解：
>
> $$\nabla_{x,\lambda} L = 0 \tag{27}$$

现在将 Lagrange multiplier 方法应用到我们的问题上，会发现新的目标函数是：

$$J = \int_{t_0}^{t_\perp} \{L + \boldsymbol{\lambda}^T(t)\,[f(\boldsymbol{x}, \boldsymbol{u}, t) - \dot{\boldsymbol{x}}]\} dt \tag{28}$$

因此可以引入一个新的标量函数 H，即 Hamiltonian：

$$H(\boldsymbol{x}, \boldsymbol{u}, t) = L(\boldsymbol{x}, \boldsymbol{u}, t) + \boldsymbol{\lambda}^T(t) f(\boldsymbol{x}, \boldsymbol{u}, t) \tag{29}$$

物理学上，$\boldsymbol{f}$ 的单位是速度，而 L 的单位是能量，所以 $\boldsymbol{\lambda}$ 应该具有 动量 的单位。

极小值原理

Lev Pontryagin (1908-1988) 提出了 极小值原理，是经典变分法的推广。经典变分法的最优条件是：

$$\frac{\partial H}{\partial \boldsymbol{u}} = \boldsymbol{0} \tag{30}$$

极小值原理将最优条件改成是：

$$\min_{u \in \Omega} H(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{u}, t) = H(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^*, t) \tag{31}$$

即是说：在最优轨迹 $\boldsymbol{x}^*(t)$ 和最优控制 $\boldsymbol{u}^*(t)$ 上，H 取最小值。它的好处是，当 $\dfrac{\partial H}{\partial \boldsymbol{u}}$ 不连续或不存在时，或者 $\boldsymbol{u}$ 受其他约束时，也可以应用。

粗略来说，极小值原理比经典变分法更一般，而动态学习又比极小值原理更一般。

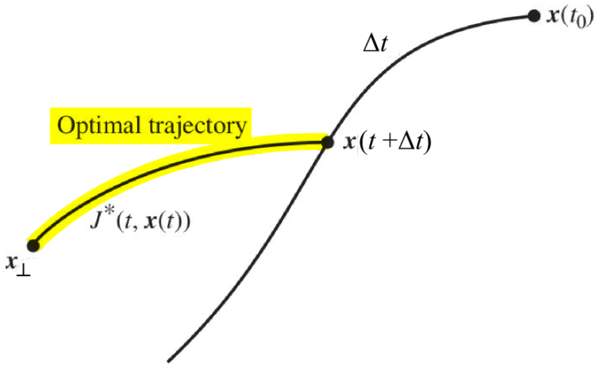
Hamilton-Jacobi-Bellman 方程

- Stanislaw Zak (2003): *Systems and control*

用动态规划的 Bellman optimality condition 可以推导出微分形式的 Hamilton-Jacobi-Bellman 方程。重温一下，Bellman 最优条件说的是：「从最优路径末端切去一小截之后，馀下的还是最优路径。」它通常写成如下的 recursive 形式：

$$\boxed{\text{最优路径}} = \text{在小段上选取最大奖励} + \boxed{\text{馀下的最优路径}} \tag{32}$$

$$J_t^* = \max_u\{\boxed{\text{奖励(u, t)}} + J_{t-1}^*\} \tag{33}$$



(34)

我们在时间 interval 的开端切出一小段：

$$[t_0, t_\perp] = [t_0, t_0 + \Delta t] \cup [t_0 + \Delta t, t_\perp] \tag{35}$$

我们想优化的目标函数是：

$$J^*(t, \boldsymbol{x}, \boldsymbol{u}) = \min_u\{\int_{t_0}^{t+\Delta t} L d\tau + \int_{t+\Delta t}^{t_\perp} L d\tau + \Phi(t_\perp, \boldsymbol{x}_\perp)\} \tag{36}$$

根据 Bellman 条件，目标函数变成：

$$J^*(t, \boldsymbol{x}) = \min_u\{\int_{t_0}^{t+\Delta t} L d\tau + J^*(t + \Delta t, \boldsymbol{x}(t + \Delta t))\} \tag{37}$$

用 Taylor series 展开右面的 J^*：

$$J^* + \frac{\partial J^*}{\partial t}\Delta t + \frac{\partial J^*}{\partial \boldsymbol{x}}(\boldsymbol{x}(t + \Delta t) - \boldsymbol{x}(t)) + \text{H.O.T.} \tag{38}$$

左右两边的 J^* 互相消去，而且 $\boldsymbol{x}(t + \Delta t) - \boldsymbol{x}(t) \approx \dot{\boldsymbol{x}}\Delta t$，於是有：

$$0 = \min_u\{\int_{t_0}^{t+\Delta t} L d\tau + \frac{\partial J^*}{\partial t}\Delta t + \frac{\partial J^*}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}}\Delta t + \text{H.O.T.}\} \tag{39}$$

又由於 Δt 很小，而且 $\dot{\boldsymbol{x}} = \boldsymbol{f}$，所以：

$$0 = \min_u\{L\Delta t + \frac{\partial J^*}{\partial t}\Delta t + \frac{\partial J^*}{\partial \boldsymbol{x}}\boldsymbol{f}\Delta t + \text{H.O.T.}\} \tag{40}$$

全式除以 Δt 并令 $\Delta t \to 0$：

$$0 = \frac{\partial J^*}{\partial t} + \min_u \{L + \frac{\partial J^*}{\partial \boldsymbol{x}} \boldsymbol{f}\} \tag{41}$$

记得 Hamiltonian 的定义是 $H = L + \dfrac{\partial J^*}{\partial \boldsymbol{x}} \boldsymbol{f}$，所以得到想要的结果：

$$\boxed{\text{Hamilton-Jacobi equation}} \quad 0 = \frac{\partial J^*}{\partial t} + \min_u H \tag{42}$$

这个方程和量子力学中的 **Schrödinger equation** 很相似：

$$i\hbar \frac{\partial}{\partial t} \Psi(x,t) = \left[V(x,t) + \frac{-\hbar^2}{2\mu} \nabla^2 \right] \Psi(x,t). \tag{43}$$

其中 Ψ 类似於我们的 J （或许 Ψ 是自然界希望取极值的某种东西？）

1.4 Symplectic 结构

- Stephanie Singer (2001): *Symmetry in mechanics – a gentle, modern introduction*
- 锺万勰 2011：《力、功、能量与辛数学》

Symplectic 的拉丁文意思是「互相交错 (intertwined)」，它用来描述 Hamiltonian 系统的几何结构。中文译作「辛」是音译。Symplectic 概念是 Hermann Weyl 研究 Hamilton 系统的对称性时在 1939 年提出的。

在数值计算上，处理 Hamilton 系统时，如果算法尊重 symplectic 结构（叫 symplectic integrators），会比一般的算法更准确；而一般解微分方程的算法，例如 Euler 算法和 Runge-Kutta 算法，有时会给出错误的结果。

举例来说，从 Hamiltonian 的角度来看，动量 p (momentum) 和速度 v (velocity) 是成对偶的，p 总是伴随 v 出现，因为 $p \cdot v = mv^2$ 的单位是能量。

举另一个例子，假设我们有两个用来定义系统状态的向量：

$$x_1 = \begin{pmatrix} s_1 \\ f_1 \end{pmatrix}, \quad x_2 = \begin{pmatrix} s_2 \\ f_2 \end{pmatrix} \tag{44}$$

其中 s 是位移（单位是长度），f 是力。这两个向量的「辛内积」定义为：

$$\begin{aligned} \langle x_1, x_2 \rangle &= x_1^T \, J \, x_2 \\ &= \begin{pmatrix} s_1 \\ f_1 \end{pmatrix}^T \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} s_2 \\ f_2 \end{pmatrix} \\ &= f_2 s_1 - f_1 s_2 \end{aligned} \tag{45}$$

其中矩阵 J 就是辛的微分形式 ω 的结构矩阵（下述）。由於 $f \cdot s$ 表示的是「所做的功」，上式表示的是

$$\text{(状态 1 的力对状态 2 的位移所做的功)} - \text{(状态 2 的力对状态 1 的位移所做的功)} \tag{46}$$

也就是「相互功」，辛正交则 $\langle x_1, x_2\rangle = 0$，代表 work reciprocity（功的互等），所以辛几何是一种关於能量的代数。

在微分几何里，研究抽象的 Hamiltonian systems，会发现 symplectic 结构。这结构用微分流形 M 及其上的一个微分形式 (differential form) ω 来定义。需要一些微分几何的基础.....

Vectors and co-vectors

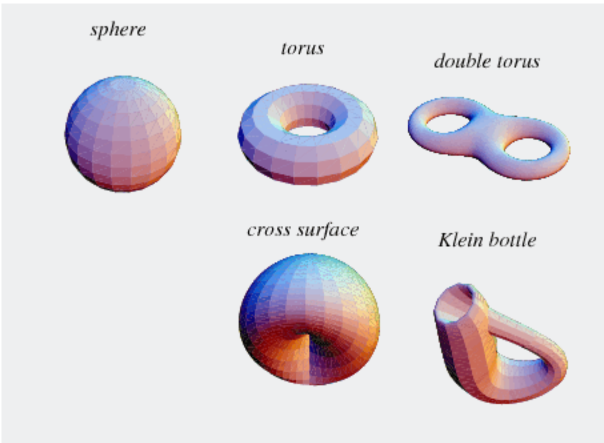
Vector 和 co-vector 之间的关系，可以看成是「d 别人的东西」和「被别人 d 的东西」，这里 d 表示微分。

「d 别人的东西」是线性的 differential operators，记作 $\dfrac{\partial}{\partial x_1}$、$\dfrac{\partial}{\partial x_2}$ 等；它们很自然地组成一个 vector space $T_x M$。

「被别人 d 的东西」是一些线性的微分形式，记作 dx_1、dx_2 等；它们属於 $T_x M$ 的 dual space。

Manifolds

基本上「流形」的意思是「弯曲的空间」，它们局部地近似於 Euclidean 空间 $\mathbb{R}^n$，局部的座标可以分段用一组微分映射来描述，这些 maps 叫 charts。



(47)

Phase space

「相位空间」指的是力学系统里，i 个粒子的位置 x_i 和动量 p_i 合并而成的 $(\boldsymbol{x}, \boldsymbol{p})$ 空间。但 configuration space 指的是所有可容许的位置 $\boldsymbol{x}$ 的空间。

Vector fields, differential forms, Hamiltonian flow

根据 Hamilton 方程，再用微分的 chain rule 可以得到：

$$\frac{d}{dt} = \frac{\partial H}{\partial \boldsymbol{p}} \frac{\partial}{\partial \boldsymbol{x}} - \frac{\partial H}{\partial \boldsymbol{x}} \frac{\partial}{\partial \boldsymbol{p}} \tag{48}$$

它是一个微分算子，亦即是向量场；它有个特别的名字叫 Hamiltonian flow $\vec{H}$ （很多书记作 X_H）：

$$\vec{H} := \frac{d}{dt} = \{\cdot, H\} \tag{49}$$

可以看出 $\vec{H}H = \{H, H\} = \frac{dH}{dt} = 0$ 就是能量守恒的形式。

Hamilton 系统的动态方程就是：

$$\dot{\boldsymbol{x}} = \vec{H} \tag{50}$$

所以在我们的智能系统中，RNN 可以看成是 $\vec{H}$。

$$\omega = \sum_i dx_i \wedge dp_i \tag{51}$$

$$\omega(\vec{H}, \cdot) = dH \tag{52}$$

我暂时不很明白它的意义。

我们说 Hamiltonian flow 保持 (preserve) 辛结构。假设 $\Gamma_t(\boldsymbol{x}_0) = \boldsymbol{x}(t)$ 描述 Hamiltonian flow 的轨迹；$\Gamma_0(\boldsymbol{x}) \equiv \boldsymbol{x}$。The pullback of ω along Γ is still ω:

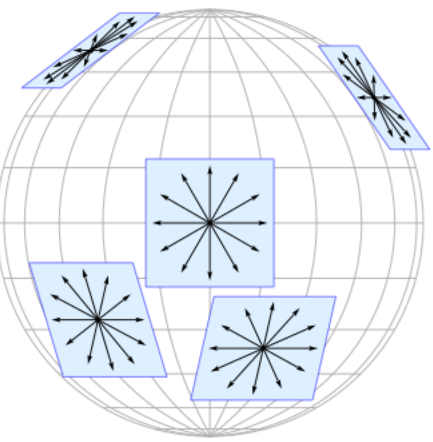
$$\Gamma_t^* \omega = \omega \tag{53}$$

$$\vec{H}H = 0 \quad \text{is equivalent to} \quad \Gamma_t^* \omega = \omega \tag{54}$$

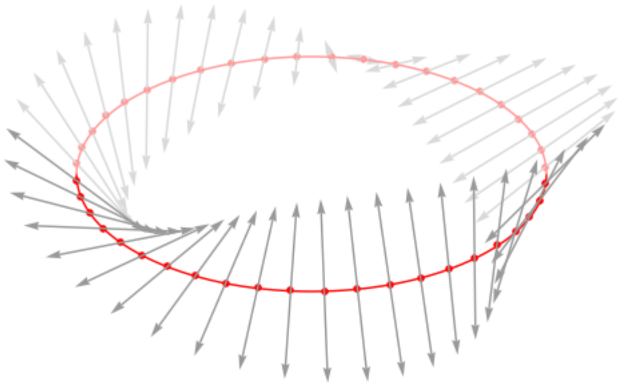
Tangent and co-tangent bundle

在流形上每点有一个 tangent space，所谓 tangent bundle 是指流形上每点 x 的 tangent space $T_x M$ 的总和：

$$TM := \bigcup_{x \in M} T_x M \tag{55}$$



Tangent bundle on a 2-sphere (56)



(57)

粗略地，tangent bundle 可以看成是 $M \times T_x M$，而 M 和 $T_x M$ 的维数都是 n，所以 tangent bundle 的维数是 $2n$。在力学上，cotangent bundle 就是相位空间 $(\boldsymbol{x}.\boldsymbol{p})$ 的空间 (The phase space of a mechanical system is the cotangent bundle of its configuration space)。

Push forward, pull back

Energy conservation, area form

(Stephanie §4.4)

Symmetry of Hamiltonian system, Lie groups

Symplectic groups 是一些保存辛结构的变换 T 的群：

$$\omega(Tx, Ty) = \omega(x, y) \quad \forall x, y \in V \tag{58}$$

V 是向量空间。在 V 上的 symplectic 变换的全体记作 $Sp(V)$。

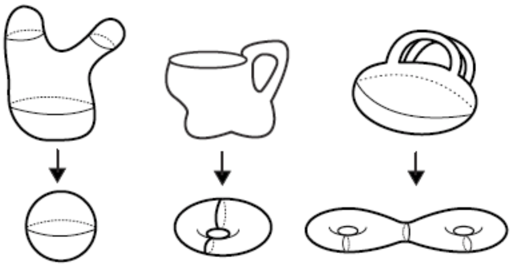
Momentum maps

1.5 动态系统理论

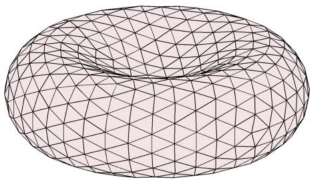
Floer homology

Homology （同调论） 研究的是空间中「有没有穿洞」的樸拓结构。最简单的 singular homology 是将空间用三角形剖分 (triangulation)，然后透过著名的 Euler formula $V + F =$

$E+2$ 及其扩充，让我们可以计算 Euler characteristic χ，此即空间穿洞的个数。

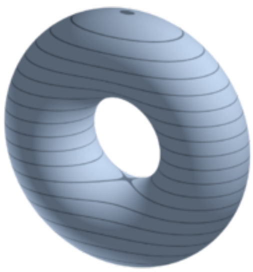


(59)



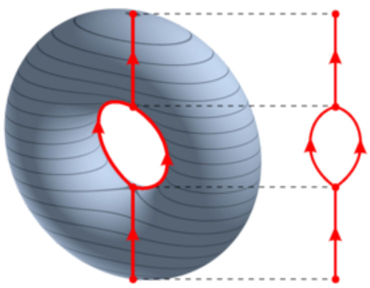
(60)

当这些三角形剖分趋於无穷小时，我们得到用微分形式 (differential forms) 描述的 homology，即 **de Rham homology**。



(61)

在流形上定义一个 potential function，例如简单的 height function，就可以做 Morse theory，这时每个点可以根据势能函数向下流 (gradient flow)，流到一些最低位置，它们是临界点 (critical points)。Morse decomposition 将空间用这些临界点分割（流到同一临界点的 flows 认作同一 equivalent class），得到的是 **Morse homology**。



(62)

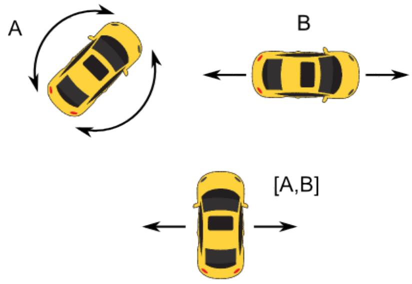
Floer homology 是 Morse homology 的**无限维空间**版本，比较难计算。

Conley theory

Entropy, ergodicity

Lie algebra 的另一应用

例如一架车可以有两个基本动作：(A) 绕中心旋转、或 (B) 前后行驶，它们的 Lie 括号 $[A, B]$，产生出的动作是左右方向行驶，这类似於「平行泊车」的时候，车子的移动方向：



(63)

（可控性与 “reachable” 概念。）

Stable, unstable, and center manifold

Smale horseshoe

2　最优控制的计算方法

直接法

间接法

Lyapunov 函数第一方法

Lyapunov 函数第二方法